Baumgartner, Meraner, Kowarik
Statistics Austria

Longitudinal Weights for the Production of Transitions and Flow Estimates

1. Introduction

In the standard course of LFS data processing, the longitudinal structure of the LFS is not in focus. Longitudinal analysis can be based on different subsamples, so one has to decide which subsample is used for weighting procedures. First, one removes all incomplete cases for the flow analysis, i.e. flows are based on the subsample of persons who are successfully surveyed in both quarters, q(t) and q(t+1). This reduction might lead to biased results and an underestimation of labour market dynamics. Second, all persons who are successfully surveyed in one quarter, first q(t) and second q(t+1) respectively, and do not regularly rotate in or out, are used. Potentially missing information of the second q(t+1) and first quarter q(t) respectively is imputed. For imputing the unobserved labour status in one quarter administrative data, e.g. from the Federation of Austrian Social Insurance Institutions and the Public Employment Service Austria, can be used. If the imputation works well, a potential bias can be reduced. Whatever subsample is used for longitudinal analysis, the cross sectional weight cannot be used. In a simple case the subsample is weighted to fit the population in the first q(t) or second quarter q(t+1). Regardless of the kind of longitudinal analysis performed, a unique identifier over all survey quarters is essential. A first step in the setting of longitudinal analysis therefore is a quality check of longitudinal data and primarily of the unique ID (in case of Austrian LFS the ID-Variable 'asbper'). The presented methodology was developed under Eurostat Grant Agreement Number 07131.2013.001-2013.371. For more detail, the final report can be requested by the authors.

2. Imputation

Preceding the weighting step, some preparations have to be made concerning missing values caused by mobile persons[1]. Important labour market characteristics are imputed via random hot deck, the standard procedure used for imputing categorical variables of Austrian LFS data (see e.g., Moser 2005), while certain variables are assumed to be stable over the respective period q(t) to q(t+1) of two quarters. The stable variables are sex, age (birth date) and other characteristics which don't change theoretically or rapidly, like migration background, nationality and highest education level.

The domain variables used for the hot deck imputation are mainly selected considering the results of the preceding bias analysis which identified a bias between mobile and immobile persons. The number of domain variables allowed for hot deck imputing Austrian LFS data is generally limited to seven. We did not change this threshold since the number of variables included in this process should not be too high because the number of donors should stay in an acceptable range. A plausibility check follows every imputation procedure. Values which are not plausible with respect to certain criteria, e.g. with respect to the labour status, are imputed again in a further imputation step. This process is repeated several times until the improvement by yet another step would be minimal. Remaining implausible values are then fixed without any stochastic effects.

All in all the shares of imputed values do not exceed 4.5% for each quarter with reference to quarter-to-quarter longitudinal datasets. Of course the year-to-year loss due to mobility is greater but does not exceed 9.6%.

[1] We call persons who were successfully surveyed in the first quarter q(t) and are lost due to panel attrition in the second quarter q(t+1) ‚outflux' or persons who move out. Persons who are only surveyed in the second quarter q(t+1) and not in the first quarter q(t) are mobile persons who move in (the sample) or 'influx'.

3. Weighting

Longitudinal plausibility checks as well as the fact of biased results and an underestimation of labour market dynamics when ignoring mobile persons lead to the usage of a subsample based upon mobile persons as well as immobile persons, the latter being persons who are successfully surveyed in both quarters, q(t) and q(t+1). Under the given circumstances, the cross sectional weight cannot be used. The adapted weighting procedure is an iterative process which considers the population at both instances.

The longitudinal weights are obtained in two main steps. First, the base weights are determined as the quotient of the population corresponding to the first quarter q(t) per NUTS-2 region and the sample size per NUTS-2 region. Thereafter, these base weights are adjusted by iterative proportional fitting to match the stock figures of certain demographic and regional characteristics. This is performed consecutively for both quarters q(t) and q(t+1). The second set of weights computed for test purposes is also calibrated against quarterly stocks regarding the labour status.

Hereinafter, the "known" marginal totals needed for fitting are derived from data of the register of residents (POPREG). The key figures for the labour status stem from published quarterly results of the microcensus, i.e. projected data.

In the following, the iterative proportional fitting procedure for the bias-reducing weighting option consists of steps 1.-2. and 4.-6. while the alternative procedure fulfilling consistency between stocks and flows consists of steps 1.-7.

1. Calibration against the distribution of the population living in private households at the beginning of the first quarter q(t) by NUTS-2 region (9 states), sex (2 groups) and age (19 groups: 1 = 0-2 years, 2 = 3-5 years, 3 = 6-9, 4 = 10-14, … (5-year classes) …, 18 = 80-84, 19 = 85+).
2. Calibration against the distribution of the population living in private households at the beginning of the first quarter q(t) by NUTS-2 region (9 states) and nationality (5 groups: 1 = Austria, 2 = EU-14, 3 = Former Yugoslavia (including also EU countries SLO and HR), 4 = Turkey, 5 = others).
3. *Calibration against the distribution of the population living in private households at the beginning of the first quarter q(t) by nationality (2 groups: 1 = Austria, 2= Not Austria), sex, age (4 groups: 1 = 0-14, 2 = 15-24, 3 = 25-64, 4 = 65+) and labour status (5 groups: 1 = employed, 2 = unemployed, 3= out of labour force, 4 = conscript in compulsory military or community service, 5 = aged younger than 15 years).*
4. Weights corresponding to people born, deceased, immigrated or emigrated in the first quarter are calibrated against the natural population change and the migration statistics.
5. Calibration like in Step 1 but against the population corresponding to the second quarter q(t+1).
6. Calibration like in Step 2 but against the population corresponding to the second quarter q(t+1).
7. *Calibration like in Step 3 but against the population corresponding to the second quarter q(t+1).*

On the one hand, it might be wise to calibrate weights to fulfil consistency between stocks and flows, especially for users. On the other hand, linking to stocks might underestimate mobile persons and important labour market dynamics, leading to biased results. Imputing for the mobile subgroup would, in such a case, not appear very useful.

For this reason, we compute two sets of weights to test both options. The first set takes advantage of imputed data and should reduce the bias whereas the second set delivers results

which are already familiar to the users. However, in view of our objective of reducing the bias concerning the labour status, it seems more promising to use the first option. It should be noted that both weighting schemes principally resemble the procedure currently used for weighting the Austrian Labour Force Survey (see e.g., Haslinger and Kytir 2006).

As an example, Table 1 to Table 3 and Figure 1 depict the difference between the two weighting versions. Weights 1 correspond to the first weighting option not adjusting for labour market status (LMS) stocks while weights 2 (LMS adj) represent the second option which does calibrate against the labour market status.

Weights 1			
	Q4 2011	**Frequency in 1000**	**in%**
	Employed	4,062.0	71.8
	Unemployed	192.3	3.4
	Out of labour force	1,403.4	24.8
	Total	5,657.7	100.0
Weights 2 (LMS adj)			
	Q4 2011	**Frequency in 1000**	**in %**
	Employed	4,089.8	72.3
	Unemployed	185.6	3.3
	Out of labour force	1,384.1	24.5
	Total	5,659.5	100.0

Table 1: Comparison of weighting options 1 and 2 for cross-sectional data Q4 2011 of population living in private households aged 15-64 without persons doing their military or civilian service according to ILO labour status .

One can see that differences are primarily noticeable for the absolute figures. Weighting scheme 1 returns a higher number of unemployed and inactive people while weighting scheme 2 results in a higher number of employed persons.

	Diff w1 w2 in 1000	**Rel diff w1 w2 in %**
Employed	-27.8	0.7
Unemployed	6.7	3.6
Out of labour force	19.3	1.4
Total	-1.8	0.0

Table 2: Differences and relative differences between the results of weighting types 1 and 2.

A comparison of the flows from both weighting procedures also shows differences with respect to the labour status:

Weights 1		**Q4 2012**			
	Q4 2011	**Frequency in 1000**			
		Employed	Unemployed	Out of labour force	Total
	Employed	3,671.7	60.9	193.1	3,925.7
	Unemployed	102.9	41.2	43.4	187.5
	Out of labour force	243.3	57.3	1,067.1	1,367.6
	Total	4,017.9	159.3	1,303.5	

Weights 2 (LMS adj)		**Q4 2012**			
	Q4 2011	**Frequency in 1000**			
		Employed	Unemployed	Out of labour force	Total
	Employed	3,720.0	79.1	210.5	4,009.6
	Unemployed	91.1	50.2	37.2	178.5
	Out of labour force	205.5	47.5	1,034.7	1,287.7
	Total	4,016.6	176.7	1,282.4	
Weights 1		**Q4 2012**			
	Q4 2011	**in %**			
		Employed	Unemployed	Out of labour force	Total
		93.5	1.6	4.9	100.0
		54.9	21.9	23.2	100.0
		17.8	4.2	78.0	100.0
Weights 2 (LMS adj)		**Q4 2012**			
	Q4 2011	**in %**			
		Employed	Unemployed	Out of labour force	Total
		92.8	2.0	5.2	100.0
		51.0	28.1	20.9	100.0
		16.0	3.7	80.4	100.0

Table 3: Comparison of weighting options 1 and 2 for flows corresponding to longitudinal data comprised of Q4 2011 and Q4 2012, i.e. the population living in private households at both time points, aged 15-64 and not doing their military or civilian service according to ILO labour status.

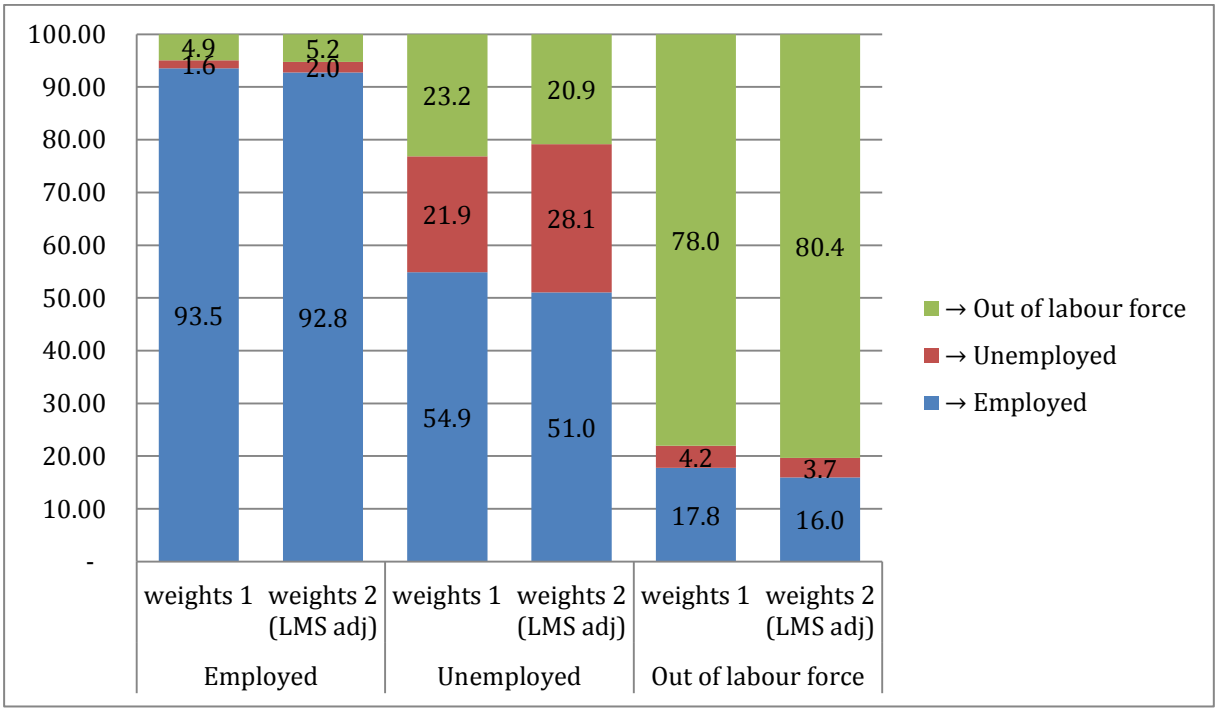

Figure 1: Comparison of weighting options 1 and 2 for flows corresponding to longitudinal data comprised of Q4 2011 and Q4 2012, i.e. the population living in private households at both time points, aged 15-64 and not doing their military or civilian service according to ILO labour status.

4. Conclusion

For producing transition rates and flow estimates we had to take several preparatory steps. We could not use the quarterly MC data set because not the whole subsample has been surveyed in both consecutive quarters. Some persons have left the sample, some have entered the sample. For producing a longitudinal data set we selected four fifth of the quarterly sample (due to rotation scheme) and did not exclude mobile persons. This led to missing values for one or another quarter due to the mobile group.

Subsequently, we checked the data for consistency and plausibility. Due to the above-mentioned circumstances, especially due to panel attrition and panel gain we could not simply use quarterly weights anymore. Missing data of the current or the previous quarter had to be estimated through imputation. In this process we had to evaluate a potential bias between immobile persons and mobile persons. Mobile persons, influx and outflux, constitute about 8% of the longitudinal data set (except for the year-to-year loss of mobile persons which is double the share of the quarter-to-quarter loss). For the moment, the quarter-to-quarter-mobiles are a small group which differs from immobile persons both in its sociodemographic structure and in its transition rates of labour market states. In the future however, this group could increase because of increasing migration flows within the Europe. Furthermore, the mobile group is much bigger for the year-to-year period. In this context we drafted two versions of longitudinal weights. The weighting option with adjustment to labour market stocks would underestimate the bias of the mobile group, therefore we decided to use the weighting option without adaption to labour market stocks for further analysis of labour market dynamics and production of flow estimates.

5. References

Haslinger, A./Kytir, J. (2006): 'Stichprobendesign, Stichprobenziehung und Hochrechnung des Mikrozensus ab 2004', Statistische Nachrichten 6/2006, 510-519, Wien.

Moser, W. (2005): 'Das Datenmanagement im neuen Mikrozensus - Eine Prozessbeschreibung', Austrian Journal of Statistics, 34 (4), 327-343, Wien.